TwoQuarks

Precision Systems for Adaptive Intelligence · Apr 2026

twoquarks.com
research@twoquarks.com

Is inference-time stability regulation sufficient to prevent collapse and unsafe behavior in sequence models under regime shift?

About TwoQuarks

TwoQuarks is an independent AI safety research project founded by Luis Jaime Ledesma (Guadalajara, México). It defines collapse in language models as a stability failure during inference — not a capacity problem. The project introduces a modular control layer that monitors pre-instability signals and applies targeted interventions at runtime, without modifying parameters, policies, or training objectives. All work ships as reproducible artifacts: probe libraries, analysis scripts, and preprints enabling systematic inspection of system behavior before deployment.

Framework — Isomeric Polarization (IP)

Isomeric Polarization (IP) is a behavioral measurement framework for LLMs under contextual pressure. It treats semantically equivalent content under different surface framings as *isomers* — inputs that may elicit divergent internal states. The core metric **DeltaL3** captures aggregate distributional shift across six quark flavors (Down, Strange, Up, Charm, Top, Bottom), each mapping a distinct failure mode: sycophancy, refusal erosion, anchor displacement, narrative override, and reasoning drift.

The Molecule pipeline operationalizes IP as a black-box probe using invariant, non-parametric metrics (TF-IDF cosine, Jaccard, n-gram overlap, length ratio) applied pairwise across response variants. **A quark that learns can be trained to tolerate drift instead of detecting it.** Molecule's metrics are fixed — no training, no weight access required.

C2 Refusal Erosion	C3 Anchor Displacement	Architectures Validated
rho = +0.713 | p = 0.0013 · Claude Haiku | 17 context depths · seed=1054	**p = 0.054** | Strongest cross-arch signal Claude · GPT-4o-mini · Mistral	**3 production LLMs** | Statistically significant regime separation (p < 0.05)

Instruments — Molecule v0.1.0

Python Package
pip install twoquarks
OpenAI, Anthropic, Ollama, HuggingFace, LiteLLM. Zero hard dependencies. Full CLI included.

Live Playground
twoquarks.com/quarkslab
Browser-based probe runner. Explorer tier open, no token required. DeltaL3 analysis with no install.

Services — Research on Your Team

Behavioral Stability Audits
Black-box probing with Molecule/PfV. Pre-collapse signals detected before production.

Inference-time Instrumentation
TwoQuarks control layer on existing pipelines. No weight modification. No retraining.

Cross-Architecture Research
Validated across Claude, GPT, Mistral. Open to funded collaborations.

RAG & Agentic Systems
Retrieval-augmented pipelines, tool-calling agents, and MCP integrations.

Azure AI & Enterprise Stack
Azure AI Foundry, Copilot Studio, multi-agent orchestration. Research rigor in production.

Custom AI Tooling
Python, PyTorch, REST APIs, evaluation frameworks — prototype to deployment.